January, 1984 LIDS-P-1351

Detection of Abrupt Changes in Dynamic Systems*

Alan S. Willsky

Dept. of Electrical Engineering and Computer Science

and

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, Massachusetts USA

Abstract

In this paper we present some of the basic ideas associated with the detection of abrupt changes in dynamic systems. Our presentation focuses on two classes of methods -- multiple filter-based techniques and residual-based methods -- and in far more detail on the multiple model and generalized likelihood ratio methods. Issues such as the effect of unknown onset time on algorithm complexity and structure and robustness to model uncertainty are discussed.

*The work described in this paper was supported in part by the U.S. Office of Naval Research Under Grant N00014-77-C-0224 and by NASA Ames and NASA Langley Research Centers under Grant NGL-22-009-124 .

1. Introduction

In recent years many techniques have been proposed for the detection of abrupt chnages in dynamic systems. These efforts have been motivated by a wide variety of applications including the detection of sensor and actuator failures [1, 2, 4, 19, 26-35] the tracking of maneuvering vehicles [20, 21, 23, 25], and numerous signal analysis problems (electrocardiogram analysis [5, 6], geophysical signal processing [7], edge detection in images [8, 9], freeway monitoring [10, 11],...). A key to the development of any technique for the detection of abrupt changes is the modeling of how the abrupt change affects the observed signals. In some applications the effect of the abrupt change is direct and simple -- e.g. a bias developing in an output signal. In such problems the primary focus of research is on the precise nature of the decision rule (see, for example [8, 9, 26]). In other applications (such as those described in [1, 2, 4, 10, 11, 19, 21]), the effect on the observables is described in a more complex, indirect way -- for example, in terms of an abrupt change in the dynamics of a system. In such problems one is presented in essence with two problems: the processing of the observed signals in order to accentuate (and simplify) the effect of the abrupt change and the definition of decision statistics and rules in terms of the processed outputs. The techniques described in this paper in principle address both of these issues in that they produce sufficient statistics for optimum detection. However, we will focus for the most part on the first task of change detection, that is, the problem of producing signals which make subsequent detection as easy as possible. As discussed here and in more detail in [27-29], this is an exceedingly important perspective in the design of detection methods which are robust to uncertain details of the dynamic models on which they are based.

In [1] a variety of methods and structures are described for change

detection. In this paper we focus on two basic and extremely important structures. The first of these is the multiple filter structure depicted in Figure 1. Here the observations, y, are processed by a bank of filters each of which is based on a particular hypothesis (e.g. Filter #1 assumes no change has occurred, Filter #2 assumes a particular type of change has occurred possibly at a particular time, etc.) The outputs of the filters, γ, represent signals which should typically be small if the corresponding hypotheses are in fact correct, and thus the decision mechanism in essence is based on determining which of the filters is doing the "best" job of keeping the corresponding γ's small. There are several methods that have been developed which fit the general form of Figure 1. In particular, hard [33] and soft [34] voting systems can be interpreted in this fashion. Another example is the multiple observer design described in [36]. In the next section we describe in detail a third technique of this general type, namely the multiple model method.

A second general structure for the detection of abrupt changes is the residual-based structure illustrated in Figure 2. In this case a filter is designed based on the assumption that no abrupt change has occurred or will occur. The filter produces a prediction $\hat{y}$ of the output signal y based on this assumption and the past history of the output, and this prediction is subtracted from the actual output to produce a residual signal γ. If no abrupt change has occurred, γ should be small. Consequently deviations from this behavior are indicative of failure, and it is on this fact that the decision mechanism is based. Again there are a variety of techniques of this general form. In [35] a variety of statistical tests (chi-squared, whiteness, etc.) are proposed for the detection of abrupt changes when the γ are the innovations from a Kalman filter. In [30-32] a method is described for the choice of gain in an observer-like filter in order to guarantee that the decoupling of the steady-state effects on γ of a given set of possible

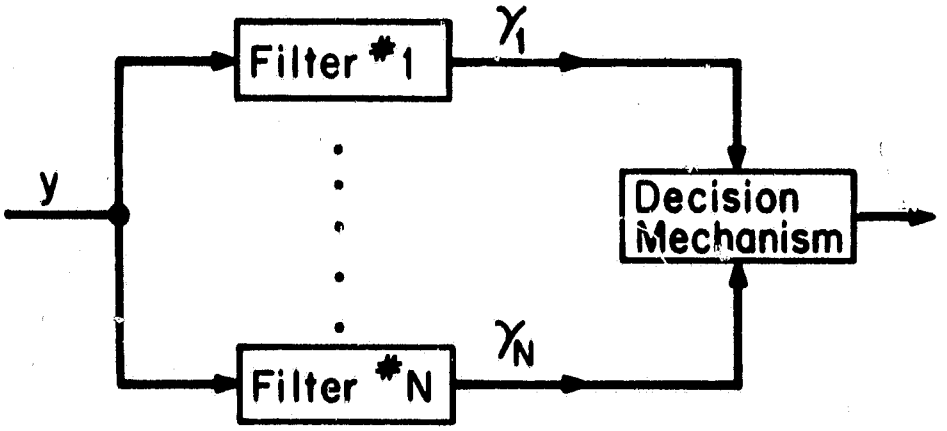

Figure 1

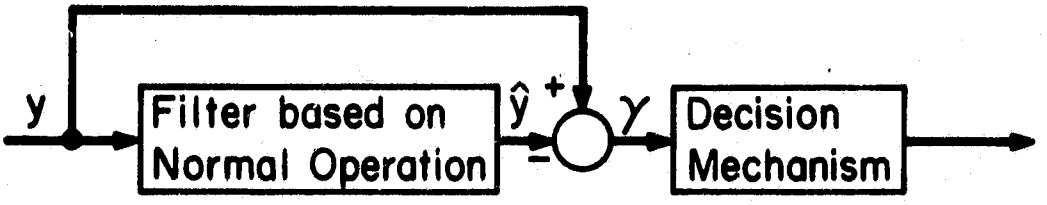

Figure 2

abrupt changes. In Section 3 we discuss a third technique of this general type, namely the generalized likelihood ratio method.

2. The Multiple Model (MM) Method

The MM method was originally developed for problems of system identification and adaptive control [12-17, 24], and in the initial part of this section we follow these early treatments. Subsequently we will look more closely at the issues that arise and possible adaptations that may be necessary for the use of MM for the detection of abrupt events (see [1, 2, 5, 10, 18, 19, 22, 23] for further developments).

The basic MM method deals with the following problem. We observe the inputs $u(k)$, $k = 0,1,2,\ldots$ and outputs $y(k)$, $k = 1,2,\ldots$ of a system which is assumed to obey one of a given finite set of linear stochastic models, indexed by $i = 1,\ldots,N$:

$$x_i(k+1) = A_i(k)x_i(k) + B_i(k)u(k) + w_i(k) + g_i(k) \tag{2.1}$$

$$y(k) = C_i(k)x_i(k) + v_i(k) + b_i(k) \tag{2.2}$$

where $w_i(k)$ and $v_i(k)$ are independent, zero-mean Gaussian white noise processes, with

$$E[w_i(k)w_i(j)'] = Q_i(k)\delta_{jk} \tag{2.3}$$

$$E[v_i(k)v_i(j)'] = R_i(k)\delta_{jk} \tag{2.4}$$

The initial state $x_i(0)$ is assumed to be Gaussian, independent of w_i and v_i, with mean $\hat{x}_i(0|0)$ and covariance $P_i(0|0)$ (the meaning of this notation will become clear in a moment). The matrices $A_i(k)$, $B_i(k)$, $C_i(k)$, $Q_i(k)$, and $R_i(k)$ are assumed to be known. Also, $b_i(k)$ and $g_i(k)$ are given deterministic functions of time (corresponding to biases, linearizations about different operating points, etc.). In addition, the state vectors $x_i(k)$ may be of different dimensions for different values of i (corresponding to assuming that the different hypothesized models represent different

order) for the dynamics of the real system). There are a number of issues that can be raised concerning this formulation, and we defer our critique of the MM method until after we have developed its basic structure. We note here only one technical point which is that we will focus on a discrete-time formulation of the MM method. Continuous-time versions can be found in the literature (see [24]), and they differ from their discrete-time counterparts only in a technical and not in a conceptual or structural manner.

Assuming that one of these N models is correct, we now have a standard multiple hypothesis testing problem. That is, let H_i denote the hypothesis that the real system corresponds to the ith model, and let $p_i(0)$ denote the a priori probability that H_i is true. Similarly, let $p_i(k)$ denote the probability that H_i is true based on measurements through the kth measurement, i.e. given $I_k = \{u(0),\ldots,u(k-1), y(1),\ldots,y(k)\}$. Then Bayes' rule yields the following recursive formula for the $p_i(k)$

$$p_i(k+1) = \frac{p(y(k+1)\,|\,H_i,I_k,u(k))\,p_i(k)}{\sum_{j=1}^{N} p(y(k+1)\,|\,H_j,I_k,u(k))\,p_j(k)} \tag{2.5}$$

Thus, the quantities that must be produced at each time are the conditional probability densities $p(y(k+1)\,|\,H_i,I_k,u(k))$ for i=1,...,N. However, conditioned on H_i, this probability density is <u>precisely</u> the one step prediction density produced by a Kalman filter based on the ith model.

That is, let $\hat{x}_i(k+1|k)$ be the one-step predicted estimate of $x_i(k+1)$ based on I_k and u(k), assuming that H_i is true. Also let $\hat{x}_i(k+1|k+1)$ denote the filtered estimate of $x_i(k+1)$ based on $I_{k+1}=\{I_k,u(k),y(k+1)\}$ and the ith model. Then these quantities are computed sequentially from the following equations:

$$\hat{x}_i(k+1|1) = A_i(k)\hat{x}_i(k|k) + B_i(k)u(k) + g_i(k) \tag{2.6}$$

$$\hat{x}_i(k+1|k+1) = \hat{x}_i(k+1|k) + K_i(k+1)\gamma_i(k+1) \tag{2.7}$$

where $\gamma_i(k+1)$ is the measurement innovations process

$$\gamma_i(k+1) = y(k+1) - C_i(k)\hat{x}_i(k+1|k) \tag{2.8}$$

and $K(k+1)$ is calculated off-line from the following set of equations:

$$P_i(k+1|k) = A_i(k)P_i(k|k)A_i'(k) + Q_i(k) \tag{2.9}$$

$$V_i(k+1) = C_i(k)P_i(k+1|k)C_i'(k) + R_i(k) \tag{2.10}$$

$$K_i(k+1) = P_i(k+1|k)C_i'(k)V_i^{-1}(k+1) \tag{2.11}$$

$$P_i(k+1|k+1) = P_i(k+1|k) - K_i(k+1)C_i(k)P_i(k+1|k) \tag{2.12}$$

Here $P_i(k+1|k)$ denotes the estimation error covarinace in the estimate $\hat{x}_i(k+1|k)$ (assuming H_i to be true), and $P_i(k+1|k+1)$ is the covariance of the error $x_i(k+1) - \hat{x}_i(k+1|k+1)$, again based on H_i. Also under hypothesis H_i, $\gamma_i(k+1)$ is zero mean with covariance $V_i(k+1)$, and it is normally distributed (since we have assumed that all noises are Gaussian). Furthermore, conditioned on H_i, I_k, and $u(k)$, $y(k+1)$ is Gaussian, has mean $C_i(k)\hat{x}_i(k+1|k)$ and covariance $V_i(k+1)$. Thus, from (2.8) we deduce that

$$p(y(k+1)|H_i,I_k,u(k)) = \frac{1}{(2\pi)^{m/2}[\det V_i(k+1)]^{1/2}} \exp\{-\frac{1}{2}\gamma_i'(k+1)V_i^{-1}(k+1) \cdot \gamma_i(k+1)\} \tag{2.13}$$

where m is the dimension of y.

Equations (2.5) - (2.8) and (2.13) define the MM algorithm. The inputs to the procedure are the $y(k)$ and $u(k)$, and the outputs are the $p_i(k)$. The implementation of the algorithm can be viewed as consisting of a bank of N Kalman filters, one based on each of the N possible models. The outputs of these Kalman filters are the innovations sequences $\gamma_i(k+1)$, which effectively measure how well each of the filters can track and predict the behavior of the observed data. Specifically, if the ith model is correct, then the

one-step prediction error $\gamma_i(k)$ should be a white sequence, resulting only from the intrinsic uncertainty in the ith model. However if the ith model is not correct, then $\gamma_i(k)$ will not be white and will include errors due to the fact that the prediction is based on an erroneous model. Thus the probability calculation (2.5), (2.13) basically provides a quantitative way in which to assess which model is most likely to be correct by comparing the performances of predictors based on these models.

Let us now address several of the most important questions that arise in understanding how the MM algorithm should be used. Clearly a very important question concerns the use of MM in problems in which the real system is nonlinear and/or the noises are non-Gaussian. The answer to this problem is extremely application-dependent. The Gaussian assumption is basically used in one place--i.e. in the evaluation of $p(y(k+1)|H_i, I_k, u(k))$ in (2.13). It has been our experience that using this formula, even when $\gamma_i(k+1)$ is non-Gaussian, causes essentially no performance degradation. As we have pointed out, what MM really attempts to do is to calculate a measure of how well each of the Kalman filters is tracking by looking at the prediction errors $\gamma_i(k+1)$, and the $p_i(k)$ are simply measure of how well each of the models are tracking relative to each other and to how well we would expect them to be tracking. The critical term in (2.13) in general is

$$\gamma_i(k+1)V_i^{-1}(k+1)\gamma_i(k+1) \tag{2.14}$$

which is the square of the tracking error normalized by the predicted covariance of these errors assuming H_i is true. Thus if this quantity is large, we would tend to disregard the ith model, while if this is small, the ith filter is tracking well. The $p_i(k)$ exhibit exactly this type of behavior, and thus we can expect MM to be reasonably robust to non-Gaussian statistics. Of course this depends upon the application, but we have had good success in several applications [5, 10] in which the noises were

decidedly non-Gaussian.

As far as the nonlinearity of the real system is concerned, an obvious approach is to linearize the system about a number of operating points for each possible model and use these linearized models to design extended Kalman filters which would be used in place of Kalman filters in the MM algorithm. Again the utility of this approach depends very much on the particular application. Essentially the issue is whether the tracking error from the extended Kalman filter corresponding to the linearized model "closest to" the true, nonlinear system is markedly smaller than the errors from filters based on "more distant" models. This is basically a signal-to-noise ratio problem, similar to that seen in the idealized MM algorithm in which everything is linear. In that case the noise is measured by the $V_i(k+1)$. The larger these are, the harder it will be to distinguish the models (the quantity in (2.14) becomes smaller as V_i is increased, and this in turn tends to flatten out (as a function of i) the probabilities in (2.13)). In the nonlinear case, the inaccuracies of the extended Kalman filters effectively increase the $V_i(k+1)$ thus reducing their tracking capabilities and making it more difficult to distinguish among them. Therefore, the performance of MM in this case will depend upon how "far apart" the different models are, as compared to how well each of the trackers tracks. The farther apart the models are, the more signal we have; the poorer the tracking performance is, the more difficult it is to distinguish among the hypotheses.

Even if the true system is linear, there is clearly the question of the utility of MM given the inevitability of discrepancies between the actual system and any of the N hypothesized models. Again this is a question of signal-to-noise ratio, but in the linear case a number of results and approaches have been developed for dealing with this problem. For example, Bram [16] has developed a precise mathematical procedure for calculating

the distance between different linear models, and he has shown that the MM procedure will converge to the model closest to the real model (i.e. $p_i(k) \to 1$ for the model nearest the true system). This can be viewed as a technique for testing the robustness of MM or as a tool that enables us to decide what models to choose. That is, if the real system is in some set of models that may be infinite or may in fact represent a continuum of models (corresponding to the precise values of certain parameters), then Baram's results can be used to decide upon a finite set of these models that span the original set and that are far enough apart so that MM can distinguish among them. For example, in adaptive flight control (reference [17]) we may be interested in determining the flight condition (operating point) of an aircraft, and we can think of using MM by hypothesizing a set of linearized models that span the flight envelope.

Let us now turn explicitly to the problem of detecting abrupt changes. In such problems one must deal with one key issue that we have not yet discussed. Specifically, in change detection we are not simply attempting to determine which of the models given in (2.1) - (2.4) is the correct one, but rather we are trying to detect a <u>shift</u> from one model to another. That is, in this case the actual system obeys a model of the form

$$x(k+1) = A(k)x(k) + B(k)u(k) + w(k) + g(k) \tag{2.15}$$

$$y(k) = C(k)x(k) + v(k) + b(k) \tag{2.16}$$

where for each k the parameters of the model correspond to one of the hypothesized models in (2.1) - (2.4), but the model <u>may change with time</u>. While this possibility is not directly taken into account in the MM method as described to this point, this algorithm often does work well in detecting shifts without any major modification to take this possiblity into account (see, for example [5, 10]. The important issue in this is the adaptability

of MM and the purposes of the particular application.

To elaborate on this, let us first note that MM will, theoretically, eventually indicate a shift from one model to another. Two things, however, must be taken into account. In the first place, we see from (2.5) that if $p_i(k)$ is small, the $p_i(k+1)$ will grow only slowly at best. In fact, in practice we have found that numerical roundoff often leads to $p_i(k)$ being set to zero if the ith model is not valid up to time k. In this case $p_i(j)$ will be zero for all $j > k$. In order to avoid this drastic effect and also the extremely sluggish response of MM to a change in models, a lower bound is usually set on the $p_i(k)$. In different applications we have found bounds from 10^{-3} down to $10^{-.5}$ to be satisfactory, with very little sensitivity to the precise value of the bound. As a second point we note that if a particular model is not correct up until time k the Kalman filter based on this model may develop large errors. If then this model becomes correct at time k, it may take a long time before the prediction errors (2.8) decrease to reflect the validity of the model. From (2.13) and (2.5) we see that this in turn means that MM may not respond to this change for some time. In practice we have found that this is not a particularly bad problem if the errors in all of the Kalman filters remain bounded even when the model on which they are based is incorrect. If a particular real system-mismatched Kalman filter combination is unstable, then there may be problems if the system switches to the model corresponding to this filter. What we have found is a workable solution to this problem is to reset the estimates of potentially divergent Kalman filters to the estimate of the most probable model, and this is done whenever the probability of possibly diverging filters falls below a threshold (such as 10^{-2}).

With these modifications MM will respond more quickly to model changes. Whether this is adequate depends upon the application. In particular, if

fast response is needed for control purposes or because additional model shifts are possible, then one may wish to consider a problem formulation that explicitly includes model switches. Furthermore, in some applications the time at which a shift occurs is exceedingly important, and in such a case one may again prefer to use such an explicit formulation, as one must in applications such as multi-object tracking [37] in which keeping track of large numbers of possibilities is crucial.

In the next section we describe one such formulation, and in the remainder of this section we indicate how the MM formulation can be modified to incorporate model changes and what the cost is for this modification. Specifically suppose that the real system does correspond to one of the models (2.1) - (2.4) for each k but that the model may change from time to time. Clearly there are several different constraints that we can place on the possible sequences of models. For example, if there are no constraints, then there are N^{k+1} possible sequences of models over the first k time steps (any of N at k=0, any of N at t=1,.,.). Such a situation arises, for example, if one assumes that the sequence of models is governed by a finite-state Markov processes. Such models have been considered by several authors. See for example [40-42] in which, in addition to considering the problem of estimation, these authors also consider the problem of identifying the transition probability matrix for the finite-state process.

On the other hand, in many problems one is interested in detecting individual abrupt changes which are sufficiently separated in time so that they can be detected and accounted for separately. In such a case it is reasonable to allow only those sequences that start with one particular model (the "normal" model) and have a single shift to any of the other models. In this case there are (kN-k+1) possible sequences up to time k -- essentially we must account for all possible failure times.

The MM solution for any such set of possible sequences of models is conceptually identical to that discussed previously, except here in principle we must design a Kalman filter for each allowable sequence of models. The residuals from these filters are then used exactly as described earlier to compute the probabilities for all hypothesized sequences. Since the number of possible sequences and thus filters grows in time, some method for pruning the tree of hypotheses is needed. For example, we can think of throwing away very unlikely models. A variety of techniques for handling such MM trees have been considered in the literature [18, 19, 37]. While this may at first glance appear to be a hopelessly complex solution to the change detection problem, this approach is not without merit. Specifically, as in [19] this approach often provides a great deal of insight. Also, the implementation of Kalman filter trees is not only within the realm of feasibility for implementation using high speed digital hardware, but it is also unavoidable in problems such as multi-object tracking.

3. The Generalized Likelihood Ratio (GLR) Method

The starting point for the GLR method is a model describing normal operation of the observed signals or of the system which generated them. Abrupt changes are then modeled as additive disturbances to this model that begin at unknown times. While there are strong similarities between the GLR and MM formulations -- indeed in many cases one can use either approach with success -- the structure of the GLR algorithm is significantly different than that for the MM technique. As just discussed for MM, we will look at the case of a single such change, the assumption being that abrupt changes are sufficiently separated to allow for individual detection and compensation. The solution to the problem just described and applications of the method can be found in [1, 3, 5, 10, 20, 21, 25]. In this section we outline the basic ideas behind the technique and discuss some of its properties.

We assume that the system under consideration can be modeled as

$$x(k+1) = A(k)x(k) + B(k)u(k) + w(k) + f_i(k,\theta)\nu \tag{3.1}$$

$$y(k) = C(k)x(k) + v(k) + g_i(k,\theta)\nu \tag{3.2}$$

where the normal model consists of these equations without the f_i and g_i terms. These terms, $f_i(k,\theta)\nu$ and $g_i(k,\theta)\nu$, represent the presence of the ith type of abrupt change, i=1,...,N. Here θ is the unknown time at which the failure occurs (so $f_i(k,\theta) = g_i(k,\theta) = 0$ for $k < \theta$), and f_i and g_i are the specified dynamic profiles of the ith charge type. For example, if f_i=0 and g_i=a vector whose components are all zero except for the jth one which equals 1 for $k \geq 0$, then this corresponds to the onset of a bias in the jth component of y. Finally, the scalar ν denotes the magnitude of the failure (e.g. the size of a bias) which we can model as known (as in MM and as in what is called simplified GLR (SGLR)) or unknown.

Assume that we design a Kalman filter based on normal operation, i.e. by neglecting f_i and g_i. From the previous section we have that this filter is given by

$$\hat{x}(k+1|k) = A(k)\hat{x}(k|k) + B(k)u(k) \tag{3.3}$$

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + K(k+1)\gamma(k+1) \tag{3.4}$$

$$\gamma(k+1) = y(k+1) - C(k)\hat{x}(k+1|k) \tag{3.5}$$

where K, P, and V are calculated as in (2.9) - (2.12). Suppose now that a type i change of size ν occurs at time θ. Then, because of the linearity of (3.1) - (3.5) we can write

$$x(k) = x_N(k) + \alpha_i(k,\theta)\nu \tag{3.6}$$

$$\hat{x}(k|k) = \hat{x}_N(k|k) + \beta_i(k,\theta)\nu \tag{3.7}$$

$$\hat{x}(k+1|k) = \hat{x}_N(k+1|k) + \mu_i(k+1,\theta)\nu \tag{3.8}$$

$$\gamma(k) = \gamma_N(k) + \rho_i(k,\theta)\nu \tag{3.9}$$

where x_N, $\hat{x}_N$, and γ_N are the responses if no abrupt change occurs, and the

other terms are the responses due solely to the abrupt change. Straightforward calculations yield recursive equations for theses quantities:

$$\alpha_i(k+1,\theta) = A(k)\alpha_i(k,\theta) + f_i(k,\theta), \quad \alpha_i(\theta,\theta) = 0 \tag{3.10}$$

$$\beta_i(k+1,\theta) = [I-K(k+1)C(k+1)]\mu_i(k+1,\theta) + K(k+1)\cdot$$
$$\cdot[C(k+1)\alpha_i(k+1,\theta) + q_i(k+1,\theta)] \quad . \tag{3.11}$$

$$\mu_i(k+1,\theta) = A(k)\beta_i(k,\theta), \quad \beta_i(\theta-1,\theta) = 0 \tag{3.12}$$

$$\rho_i(k,\theta) = C(k)[\alpha_i(k,\theta) - \mu_i(k,\theta)] + q_i(k,\theta) \tag{3.13}$$

The important point about these quantities is that they can be precomputed. Furthermore, by its definition, $\gamma_N(k)$ is the innovations under normal conditions, i.e. it is zero-mean, white, Gaussian with covariance $V(k)$. Thus we now have a standard detection problem in white noise: we observe the filter residuals $\gamma(k)$, which can be modeled as in (3.9), and we want to detect the presence of a change (i.e. that $k \geq \theta$) and perhaps determine its identity i and estimate its time of occurrence θ and size ν, if the latter is modeled as being unknown. The solution to this problem involves matched filtering operations. First, define the precomputable quantities

$$a(k,\theta,i) = \sum_{j=\theta}^{k} \rho_i'(j,\theta)V^{-1}(j)\rho_i(j,\theta) \tag{3.14}$$

This has the interpretation as the amount of information present in $y(\theta),..,y(k)$ about a type i change occurring at time θ.

The on-line GLR calculations consist of the calculation of

$$d(k,\theta,i) = \sum_{j=\theta}^{k} \rho_i'(j,\theta)V^{-1}(j)\gamma(j) \tag{3.15}$$

which are essentially correlations of the observed residuals with the abrupt change signatures $\rho_i(j,\theta)$ for different hypothesized types, i, and times, θ. If ν is known (the SGLR case), then the likelihood of a type i change having occurred at time θ given data $y(1),...,y(k)$ is

$$\ell_s(k,\theta,i) = 2\nu d(k,\theta,i) - \nu^2 a(k,\theta,i) \tag{3.16}$$

If ν is unknown, then the generalized likelihood for this change is

$$\ell(k,\theta,i) = \frac{d^2(k,\theta,i)}{a(k,\theta,i)} \tag{3.17}$$

and the maximum likelihood estimate of ν assuming a change of type i at time θ is

$$\hat{\nu}(k,\theta,i) = \frac{d(k,\theta,i)}{a(k,\theta,i)} \tag{3.18}$$

Thus the GLR algorithm consists of the single Kalman filter (3.3) - (3.5), the matched filter operations of (3.15), and the likelihood calculation of (3.16) or (3.17). The outputs of the method are these likelihoods and the estimates of eq. (3.18) if ν is modeled as unknown. The basic idea behind GLR is that different types of abrupt changes produce different kinds of effects on the filter innovations -- i.e. different signatures -- and GLR calculates the likelihood of each possible event by correlating the innovations with the corresponding signature.

As with the MM method a number of issues can be raised about GLR. Some of these, such as the effect of nonlinearities and robustness to model errors, are very similar to the MM case. Essentially it still can be viewed as a signal-to-noise ratio problem: in the nonlinear case the additive decomposition of (3.9) is not precisely valid, but it may be approximately correct. Also, different failure modes can be distinguished even in the presence of modelling errors if their signatures are different enough. Again these issues depend very much on the particular application. We refer the reader to [4, 6, 10, 11, 21, 25] for discussions of several applications of GLR to applications in which these issues had to be addressed.

GLR has been successfully applied to a wide variety of applications, such as failure detection [1, 4], geophysical signal analysis [7], detecting arrhythmias in electrocardiograms [6], freeway incident detection [10, 11],

and maneuver detection [20, 21, 25]. Note that the model used in (3.1), (3.2) for such changes is an additive model. Thus it appears on the surface that the types of abrupt changes that can be detected by GLR are a special subset of those that can be detected by MM, since (2.1), (2.2) allow parametric changes (in A, B, C, Q, R) as well as additive ones. There are several points, however, that must be taken into account in assessing and comparing MM and GLR:

(1) The price one pays for allowing parametric changes in MM is the necessity of implementing banks of Kalman filters, and actually trees of such filters to account for switches between models. GLR, on the other hand, requires a single Kalman filter and a growing number of correlation calculations as in (3.15), which in principle must be calculated for $i=1,\ldots,N$ and $\theta=1,\ldots,k$. We will comment shortly on the computational issues concerned with these correlations, but for now we simply point out that they are typically far less involved than the calculations inherent in Kalman filters (see [4, 6, 7] for examples of how simple these calculations can be). Also, because it operates on the outputs of a normal mode filter, GLR can be easily implemented and attached as a monitor to an already existing system.

(2) Extensions to the GLR method can be developed for the detection of parametric changes [38]. This extended GLR bears some similarity to extended Kalman filtering and iterated extended Kalman filtering.

(3) It has been our experience that a GLR system based on the detection of additive effects can often also detect parameter failures. For example, a gain change in a sensor does look like a sensor bias, albeit one that is modulated by the value of the variable being sensed. That is, any detectable change will exhibit a systematic

deviation between what is observed and what is predicted to be observed. Obviously, the ability of GLR to detect a parametric change when it is looking for additive ones is again a question of robustness. If the effect of the parametric change is "close enough" to that of the additive one, the system will work. This has been the case in all of our experience. In particular we refer the reader to [4] for an additive-failure-based design that has done extremely well in detecting gain changes in sensors. Note of course that in this mode GLR is essentially only indicating an alarm -- i.e. the estimate $\hat{\nu}$ of the "bias" is meaningless, but in many detection problems our primary interest is in simply identifying which of several types of changes has occurred.

There are several final issues that should be mentioned in discussing GLR. The first concerns the calculation of statistical measures of performance of GLR. As mentioned in the preceding section, Baram [16] has developed a method for measuring the distance between models and hence a measure of the detectability and distinguishability of different failure modes. Similar calculations can be performed for GLR, but in this case it is actually simpler to do and interpret, as we can use standard detection-theoretic ideas. Specifically, a direct measure of the detectability of a particular type of change is the information $a(k,\theta,i)$ defined in (3.14). This quantity can be viewed as the correlation of $\rho_i(j,\theta)$ with itself at zero lag. Similarly, we can determine the relative distinguishability of a type i change at two times θ_1 and θ_2 as the correlation of the corresponding signatures

$$a(k,\theta_1,\theta_2,i) = \sum_{j=\max(\theta_1,\theta_2)}^{k} \rho_i'(j,\theta_1)V^{-1}(j)\rho_i(j,\theta_2) \qquad (3.19)$$

and the relative distinguishability of type i and m changes at times θ_1 and θ_2 similarly:

$$a(k,\theta_1,\theta_2,i,m) = \sum_{j=\max(\theta_1,\theta_2)}^{k} \rho_i'(j,\theta_1)V^{-1}(j)\rho_m(j,\theta_2) \qquad (3.20)$$

These quantities provide us with extremely useful information. For example, in some applications [6-9] the estimation of the time θ at which the change occurs is critical, and (3.19) provides information about how well one can resolve the onset time. In failure detection applications these quantities directly provide us with information about how system redundancy is used to detect and distinguish failures and can be used in deciding whether additional redundancy (e.g. more sensors) are needed. Also, the quantities in (3.14), (3.19), and (3.20) directly give the statistics of the likelihood measures (3.16), (3.17). For the SGLR case of (3.16), ℓ_s is Gaussian, and its mean under no failure is $-\nu^2 a(k,\theta,i)$, while if a type m failure occurs at time ϕ, its mean is

$$E[\ell_2(k,\theta,i) \,|\, (m,\phi)] = \nu^2[2a(k,\theta,\phi,i,m) - a(k,\theta,i)] \qquad (3.21)$$

For example if $(m,\ell) = (i,\theta)$ -- i.e. if the precise failure and time assumed in the calculation of $\ell_s(k,\theta,i)$ are true, then its mean is $+\nu^2 a(k,\theta,i)$. In the case of (3.17), under no failure $\ell(k,\theta,i)$ is a chi-squared random variable with 1 degree of freedom, while if a failure (m,ϕ) of size ν occurs $\ell(k,\theta,i)$ in non-central chi-squared with mean

$$E[\ell(k,\theta,i) \,|\, (m,\phi)] = 1 + \frac{\nu^2 a(k,\theta,\phi,i,m)^2}{a(k,\theta,i)} \qquad (3.22)$$

Clearly these quantities can be very useful in evaluating the performance of GLR detection algorithms and for determining decision rules based on the GLR outputs. If one were to follow the precise GLR philosophy [39], the decision rule one would use is to choose at each time k the largest of the $\ell_s(k,\theta,i)$ or $\ell(k,\theta,i)$ over all possible change types i and onset times θ.

This largest value would then be compared to a threshold for change detection, and if the threshold is exceeded the corresponding maximizing values of θ and i are taken as the estimates of change type and time. While such a simple rule works in some cases [6, 21], it is worthwhile often to consider more complex rules based on the ℓ's. For example, persistance tests (i.e. ℓ must exceed the threshold over some time period) are often used to cut down on false alarms due to spurious and unmodeled events. See [4, 7, 9, 26] for more discussion of decision rules.

A final issue to be mentioned is the pruning of the tree of possibilities. As in the MM case in principle we have a growing number of calculations to perform, as $d(k,0,i)$ must be calculated for $i=1,\ldots,N$ and all possible change times up to the present, i.e. $\theta=1,\ldots,k$. What is usually done is to look only over a sliding window of possible times:

$$k-M_1 \leq \theta \leq k-M_2 \tag{3.23}$$

where M_1 and M_2 are chosen based on the a's -- i.e. on detectability and distinguishability considerations. Basically after M_2 times steps from the onset of change we have collected enough information so that we may make a detection with a reasonable amount of accuracy. Further, after M_1 time steps we will have collected a sufficient amount of information so that detection performance is as good as it can be (i.e. there is no point in waiting any longer). Clearly we want M_1, M_2 large to allow for maximum information collection, but we want them small for fast response and for computational simplicity. This is a typical tradeoff that arises in all change detection problems.

REFERENCES

1. Willsky, A.S., "A Survey of Several Failure Detection Methods," Automatica, Nov. 1976, pp. 601-611.

2. Willsky, A.S., "Failure Detection in Dynamic Systems," AGARD Lecture Series No. 109 on Fault Tolerance Design and Redundancy Management Technqiues, Athens, Rome, and London, Oct. 1980.

3. Willsky, A.S., and Jones, H.L., "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems," IEEE Trans. Automatic Control, Vol. AC-21, No. 5, Feb. 1976, pp. 108-112.

4. Deckert, J.C., Desai, M.N., Deyst, J.J. and Willsky, A.S., "F8-DFBW Sensor Failure Identification Using Analytic Redundancy," IEEE Trans. on Automatic Control, Vol. AC-22, No. 5, Oct. 1977, pp. 795-803.

5. Gustafson, D.E., Willsky, A.S., Wang, J.-Y., Lancaster, M.C., and Triebwasser, J.H., "ECG/VCG Rhythm Diagnosis Using Statistical Signal Analysis I: Identification of Persistent Rhythms," IEEE Trans. Biomed. Eng., Vol. BME-25, No. 4, July 1978, pp. 344-353.

6. Gustafson, D.E., Willsky, A.S., Wang, J.-Y., Lancaster, M.C., and Triebwasser, J.H., "ECG/VCG Rhythm Diagnosis Using Statistical Signal Analysis II: Identification of Transient Rhythms," IEEE Trans. Biomed. Eng., Vol. BME-25, No. 4, July 1978, pp. 353-361.

7 Basseville, M. and Benvensite, A., "Design and Comparative Study of Some Sequential Jump Detection Algorithms for Digital Signals," IEEE Trans. Acous., Speech, Sig. Proc., Vol. ASSP-31, June 1983, pp. 521-534.

8. Basseville, M., Espiau, B., and Gasnier, J., "Edge Detection Using Sequential Methods for Change in Level, Part I: A Sequential Edge Detection Algorithm," IEEE Trans. Acous., Speech, Signal Processing, Vol. ASSP-29, No. 1, Feb. 1981, pp. 24-31.

9. Basseville, M., "Edge Detection Using Sequential Methods for Change in Level, Part II: Sequential Detection of Change in Mean," IEEE Trans. Acous., Speech, Signal Processing, Vol. ASSP-29, No. 1, Feb. 1981, pp. 32-50.

10. Willsky, A.S., Chow, E.Y., Gershwin, S.B., Greene, C.S., Houpt, P.K., and Kurkjian, A.L., "Dynamic Model-Based Techniques for the Detection of Incidents on Freeways," IEEE Trans. Automatic Control, Vol. AC-25, June 1980.

11. Kurkjian, A.L., Gershwin, S.B., Houpt, P.K., Willsky, A.S., Greene, C.S., and Chow, E.Y., "Estimation of Roadway Traffic Density on Freeways Using Presence Detector Data," Transportation Sciences, 1980.

12. Lainiotis, D.G., "Joint Detection, Estimation, and System Identification," Information and Control, Vol. 19, Aug. 1971, pp. 75-92.

13. Willner, D., Observation and Control of Partially Unknown Systems, Ph.D., Thesis, MIT, May 1973.

14. Magill, D.T., "Optimal Adaptive Estimation of Sampled Processes," IEEE Trans. on Automatic Control, Vol. AC-10, pp. 434-439, Oct. 1965.

15. Lainiotis, D.G., "Partitioning: A Unifying Framework for Adaptive Systems, I: Estimation," Proc. of IEEE, Vol. 64, No. 8, pp. 1126-1142, 1976.

16. Baram, Y., Information, Consistent Estimation and Dynamic System Identification, Ph.D. Thesis, MIT, Nov. 1976.

17. Athans, M., et. al., "The Stochastic Control of the F-8C Aircraft Using a Multiple Model Adaptive Control (MMAC) Method Part I: Equilibrium Flight", IEEE Trans. on Automatic Control, Vol. AC-22, No. 5, pp. 768-780, 1977.

18. Buxbaum, P.J. and Haddad, R.A., "Recursive Optimal Estimation for a Class of Nongaussian Processes," Proc. Symp. on Computer Processing in Communications, Polytech. Inst. of Brooklyn, April 8-10, 1969.

19. Willsky, A.S., Deyst, J.J., and Crawford, B.S., "Two Self-Test Methods Applied to an Inertial System Problem," J. Spacecr. Rockets, Vol. 12, No. 7, July 1975, pp. 434-437.

20. McAulay, R.J. and Denlinger, E., "A Decision-Directed Adaptive Tracker," IEEE Trans. on Aero. and Elec. Sys., Vol. AES-9, March 1973, pp. 229-236.

21. Dowdle, J.R., Willsky, A.S., and Gully, S.W., "Nonlinear Generalized Likelihood Ratio Algorithms for Maneuver Detection and Estimation," Proc. of 1982 American Control Conf; Arlington, Virginia, June 1983.

22. Newbold, P.M. and Ho, Y.C. "Detection of Changes in the Characteristics of a Gauss-Markov Process," IEEE Trans. Aerospace Elec. Sys., Vol. AES-4, No. 5, Sept. 1968, pp. 707-718.

23. Athans, M., Whiting, R.H., and Gruber, M., "A Suboptimal Estimation Algorithm with Probabilistic Editing for False Measurements with Applications to Target Tracking with Wake Phenomena," IEEE Trans. Aut. Control, Vol. AC-22, June 1977, pp. 372-384.

24. Greene, C.S., "An Analysis of the Multiple Model Adpative Control Algorithm," Ph.D. Dissertation, Report No. ESL-TH-843, M.I.T., Elec. Sys. Lab., Cambridge, Mass., August 1978.

25. Tenney, R.R., Hebbert, R.S., and Sandell, N.R., "A Tracking Filter for Maneuvering Sources," IEEE Trans. Aut. Control, Vol. AC-22, April 1977, pp. 246-251.

26. Basseville, M., "Contribution a la Détection Séquentielle de Ruptures de Modeles Statistiques," Thèse d'Etat, Univ. de Rennes, France, June 1982.

27. Chow, E.Y., and Willsky, A.S., "Analytical Redundancy and the Design of Robust Failure Detection Systems," IEEE Trans. Aut. Control, to appear.

28. Lou, X.-C., Willsky, A.S., and Verghese, G.C., "Failure Detection with Uncertain Models," Proc. 1983 American Control Conference, San Francisco, Calif., June 1983.

29. Lou, X.-C., Willsky, A.S., and Verghese, G.C., "Optimally Robust Redundancy Relations for Failure Detection in Uncertain Systems," M.I.T. Lab. for Inf. and Dec. Sys. Rept. LIDS-P-1297, April 1983; submitted to Automatica.

30. Beard, R.V., "Failure Accommodation in Linear Systems Through Self-Reorganization," Rept. MVT-71-1, Man Vehicle Laboratory, M.I.T., Cambridge, Mass. Feb. 1971.

31. Jones, H.L., "Failure Detection in Linear Systems," Ph.D. thesis, Dept. of Aero. and Astro., M.I.T., Cambridge, Mass., Sept. 1973.

32. Meserole, J.S., "Detection Filters for Fault-Tolerant Control of Turbofan Engines," Ph.D. thesis, Dept. of Aero. and Astro., M.I.T., Cambridge, MA., June 1981.

33. Gilmore, J. and McKern, R., "A Redundant Strapdown Inertial System Mechanization -- SIRU," presented at the AIAA Guidance, Control, and Flight Mechanics Conf., Santa Barbara, Calif., Aug. 1970.

34. Broen, R.B., "A Nonlinear Voter - Estimator for Redundant Systems," Proc. 1974 IEEE Conf. on Dec. and Control, Phoenix, Arizona, pp. 743-748.

35. Mehra, R.K. and Peschon, J., "An Innovations Approach to Fault Detection in Dynamic Systems," Automatica, Vol. 7, 1971, pp. 637-640.

36. Clark, R.N., Fosth, D.C., and Walton, V.M., "Detecting Instrument Malfunctions in Control Systems," IEEE Trans. Aero. and Elec. Sys., Vol. AES-11, No. 4, July 1975, pp. 465-473.

37. Keverian, K. and Sandell, N.R., "Multiobject Tracking by Adaptive Hypothesis Testing," Rept. LIDS-R-959, Lab. for Inf. and Dec. Sys., M.I.T., Cambridge, Mass., Dec. 1979.

38. Willsky, A.S., "Status Report Number One on the Development of a Methodology for the Detection of System Failure and for the Design of Fault-Tolerant Control Systems," Rept. ESL-SR-781, Lab. for Inf. and Dec. Sys., M.I.T., Cambridge, Mass. Nov. 1977.

39. Van Trees, H.L., Detection, Estimation, and Modulation Theory, Part I, John Wiley and Sons, Inc., New York, 1971.

40. Tugnait, J.K. and Haddad, A.H., "A Detection-Estimation Scheme for State Estimation in Switching Environments," Automatica, Vol. 15, July 1979, pp. 477-481.

41. Tugnait, J.K., "Detection and Estimation for Abruptly Changing Systems," Proc. 20th IEEE Conf. on Decision and Control, San Diego, Calif., Dec. 1981, pp. 1357-1362.

42. Tugnait, J.K., "Adaptive Estimation and Identification for Discrete Systems with Markov Jump Parameters," IEEE Trans. Aut. Control, Vol. AC-27, No. 6, Oct. 1982, pp. 1054-1065.